# THE EFFECT OF POLITICAL CONSOLIDATION ON LOCAL DEMOCRACY:

# EVIDENCE FROM LARGE-SCALE MUNICIPAL REFORM[§]

David Dreyer Lassen
Department of Economics
University of Copenhagen

Søren Serritzlew
Department of Political Science
University of Aarhus

February 18, 2009

Abstract:

Optimal jurisdiction size is a cornerstone of government design. Well-developed theoretical work identifies a number of factors influencing the size trade-off, and key among them is the quality of local democracy, hypothesized to be higher in more decentralized settings. However, existing studies of jurisdiction size and democracy, mostly cross-sectional in nature, yield ambiguous results due to sorting effects and problems of endogeneity. In contrast, we identify a natural experiment, a large-scale municipal reform in Denmark, which allows us to estimate a causal effect of jurisdiction size on the quality of local democracy. The reform, affecting some, but not all, municipalities, was implemented by the central government, and resulted in exogenous, and substantial, changes in municipal population size. Based on repeated survey data collected before and after the reform, we find, using various differences-in-differences and matching estimators, that jurisdiction size has a causal detrimental effect on citizens' sense of internal political efficacy.

Keywords: political participation, local democracy, natural experiment, matching, differences-in-differences, municipal politics, jurisdictional reform, political efficacy, decentralization, Tiebout

## I. Introduction

How government is structured horizontally and vertically matters for economic and political outcomes. Since Aristotle, political philosophers, political scientists and economists have debated the optimal size of a political jurisdiction. Most recently, decentralization as a solution to problems of economic and political development has been given substantial support by the World Bank (e.g. World Bank, 2000). Across the world, local, regional and central governments are constantly (re-)assessing the optimal balance between (more) centralization and (more) decentralization.

Designing the optimal structure or architecture of government resolves around a key trade-off:[1] Larger, and thereby fewer, jurisdictions allow governments to take advantage of economies of scale in local public goods production. If such production involves fixed costs, such as fixed administrative or bureaucratic costs, having more jurisdictions entails duplicating such fixed costs, which is inoptimal from a societal point of view. At the same time, more jurisdictions result in greater inter-jurisdictional spill-overs in the form of externalities, arising when governments do not take into account the effects of their decisions on other entities of government.

On the other hand, having a larger number of – on average smaller – sub-national political units also confers several advantages. The most celebrated advantage of decentralization is that it can improve service delivery of local public goods.[2] This can take place through a combination of two factors. First, having several municipalities improves allocative efficiency or preference-matching by allowing citizens to sort into different

---

[1] Treisman (2007, pp. 11-15) provides a detailed discussion of these core arguments for or against decentralization. The rest of his book expands on these arguments and provides an in-depth, critical study of the economics and politics of decentralization. Oates (1972) provides the first full specification of the problem of the optimal size of jurisdictions, trading off economies of scale with heterogeneity of preferences. Hooghe and Marks (2003) provide an overview of different types of multi-level governance.

[2] The existing empirical evidence is mixed. Faguet (2004) provides an overview and an empirical illustration from Bolivia.

jurisdictions based (in part or in whole) on their preferences over local public goods; this is the classical Tiebout (1956)-argument, applied and evaluated in a voluminous literature.[3] Second, having a more decentralized structure is also thought to improve productive efficiency (Barankay and Lockwood, 2008). This can happen through a number of channels: Smaller jurisdictions can increase electoral control and accountability (e.g. Seabright, 1996; Hindricks and Lockwood, 2005), and having several jurisdictions can, by increasing the competition for tax payers and firms, lead to less wasteful government spending, partly by encouraging yard-stick competition on the side of citizens and firms (e.g. Bordignon et al., 2004). Lastly, smaller jurisdictions are often argued to increase the quality of local democracy by promoting knowledge of local political issues and increasing local political efficacy and empowerment, leading to greater political participation (Dahl and Tufte, 1973).[4]

It is this last point which is the topic of our analysis. Political efficacy, which will be the main subject of analysis in the present paper, obviously is important in its own right if participatory democracy is valuable in itself (Almond and Verba, 1989: 136-139). Furthermore, however, the argument that decentralization improves service delivery is also critically dependent on a well-functioning local democracy, for two reasons. First, political participation has representational consequences: preference-matching, or allocative efficiency, will improve only if citizens make their preferences known through the local political process. Second, participatory democracy is a prerequisite for increasing electoral control and accountability, the two factors argued to be necessary for decentralization to improve productive efficiency.

---

[3] See, e.g., Ellickson (1971), Oates (1972), Alesina and Spolaore (1997, 2003) for formal analyses in the political economy tradition. A large literature, surveyed and debated in Lyons and Lowery (1989) and Ross and Yinger (1999), focuses specifically on metropolitan government in the US. The empirical literature on Tiebout-sorting is far from settled: Rohde and Strumpf (2003) and Banzhaf and Walsh (2008) provide evidence, respectively, against sorting based on income, but for sorting based on preferences for local public goods

[4] Often, debate on decentralization also includes regional or local identities; see Meguid (2007) for an analysis of decentralization in Europe. She finds voter response to decentralization in the UK to be determined in part by regional identities or the lack thereof.

The purpose of this paper is to evaluate empirically the claim that jurisdiction size matters in a causal sense for the quality of local democracy. While the literature on this topic, which will be reviewed in detail below, is extensive, it is almost exclusively cross-sectional in nature, as time variation in jurisdiction structure rarely is observed, and when observed is typically due to an endogenous process where (de)centralization is the outcome of a referendum or political decisions at the local level. This endogeneity, together with the fact that cross-sectional estimates may reflect the deliberate decisions of people sorting themselves into jurisdictions with different characteristics, makes it difficult to interpret existing empirical estimates as reflecting a causal effect of jurisdiction size on various dimensions of local democracy.

To get at estimates of causal effects in this setting, we turn to administrative reform. As noted by Campbell (1969), administrative reforms occasionally provide researchers in the social sciences with opportunities for estimating causal effects by generating natural experiments that exogenously determines, or at least significantly influences, the treatment assignment in a non-experimental setting (Lassen, 2005).

We exploit a recent large-scale administrative reform of the Danish municipal structure, the so-called Structural Reform (SR). As a part of the reform, individuals residing in reform (treatment) municipalities suddenly experienced, through amalgamation with one or more neighbouring municipalities, a dramatic increase in the population size of their jurisdiction, while those residing in non-reform (control) municipalities experienced no change in jurisdiction size. The reform was, as we shall argue in detail below, largely exogenous from the viewpoint of individual municipalities and definitely so for individual citizens. We investigate the consequences of this exogenous shock to jurisdiction size for one key component of democracy, that of internal political efficacy, a common, well-established measure of democracy that has been used extensively both across time and across different countries and levels of government. Our empirical analysis focuses on the short-run effects of reform, in an attempt to identify the pure effect of change in size, and is based on two individual level

4

surveys. The first was carried out in 2001, well before any mention of the reform, the second in

December 2007 and January 2008, almost a year after the actual implementation of the newly

formed municipalities, and three years after the first municipal election.

The reform process was initiated by the central government, and had the explicit

aim of generating larger municipalities to reap economies of scale. It is the fact that the reform

was imposed on municipalities by the central government combined with the fact not all

municipalities were affected by the reform in the same way that makes it an attractive setting for

studying the causal effect of jurisdiction size on the quality of local democracy. The fact that 32

municipalities were not amalgamated enables us to control for national level trends in overall

survey response as well as local political interest, which could well be affected due to both the

public debate on the large-scale reform and the concurrent revision of the system of

intergovernmental grants.[5]

We estimate the causal effect of jurisdiction size on local political efficacy using

survey data on local political efficacy collected before and after the reform in both unaffected

(control) and affected (treatment) municipalities. We find, using various differences-in-

differences and matching estimators on both repeated cross-sectional data and retrospective

evaluations, that citizens in amalgamated municipalities report lower political efficacy after the

reform, and that the decline is larger, the larger is the change in population size as a result of the

reform. The result that population size has a causal, significant effect on internal political

efficacy is robust across samples and estimators, and we demonstrate that local variations in the

amalgamation process as well as changes in local public finances and municipal political control

following reform does not affect this relationship.

The paper is structured as follows: The next section provides a theoretical

framework for thinking about size, political participation, and selection. Section three describes

---

[5] At the same time, the reform did not affect the number or the placement of polling places, or other parts
of the electoral process, which allows us to isolate population size effects on voter turnout from spatial
effects (Brady and McNulty, 2004).

the Danish structural reform, section four the data and section five lays out our research design and empirical methodology. Section six presents our results and section seven concludes with a discussion.

## II. Jurisdiction size and local democracy

The question of how jurisdiction size affects the democratic process is one of the oldest in political science. It can be traced back almost 2,500 years, to Plato and Aristotle who preferred smaller entities, large enough to be self-sufficient, but small enough to ensure that citizens can know one another's characters (Dahl and Tufte, 1973: 4-5). Plato and Aristotle did not have modern democracies in mind, but the idea that jurisdiction size affects the opportunities of citizens for participating in the democratic process is still relevant (Treisman, 2007).

In their seminal, modern treatment of the relation between size and democracy, Dahl and Tufte (1973: 13-16) mention eight channels through which jurisdiction or community size can affect patterns of political participation and efficacy: opportunities for citizens to participate, voluntary compliance and the prevalence of common norms and values, homogeneity, conceptions of the common good, loyalty, emotional ties to society, possibilities for accurate communication, and leader responsiveness to citizens. The literature has less to say about what exactly happens in smaller jurisdictions that encourages participation and feelings of efficacy; but see Oliver (2000) for an exception. We remain agnostic about this for the time being and return to the issue in the conclusion

Many empirical studies have investigated the effects of jurisdiction size and consolidation, some with specific focus on size and various dimensions of democracy, some as part of studies of decentralization more broadly.[6] Morlan (1984) argues that turnout is inversely

---

[6] A related literature, also surveyed in Treisman (2007), considers the relationship between country-level measures of (fiscal) decentralization and broad, expert-based measures of democracy. While we do not make any comparative claims, we see our present study as beginning to provide some micro-foundations for the arguments evaluated in that literature.

related to jurisdiction size, as turnout is higher in national elections than in local elections and higher in smaller municipalities, and investigates the longitudinal relationship between aggregate turnout and waves of municipal consolidation. Also other aspects of participation than turnout are affected by jurisdiction size (Verba and Nie, 1972). For instance, Oliver (2000, 2001) finds, in an analysis of a cross-section of US metropolitan jurisdictions, that voters in smaller jurisdictions are more likely to contact officials and to attend community or organizational meetings. Meguid (2007) considers devolution to regional parliaments and its effects on voter engagement. She finds no overall effects, but, comparing voter sentiments before and after the establishment of the Scottish parliament, she finds that Scottish nationalist voters report greater political engagement while no such change is observed for other voters, presumably less interested in Scottish home rule. Generally, high turnout and rates of participation is an indication that citizens find it worthwhile to engage in politics, and is related to the perception of citizens of their own competence to participate in politics. This perception, political efficacy, is correlated with the tendency to participate in politics (Stenner-Day and Fischle, 1992).

The concept of political efficacy, which will form the basis of our empirical investigation, was introduced by Campbell, Gurin and Miller (1954). They define it as the "feeling that individual political action does have, or can have, an impact upon the political process, i.e., that it is worthwhile to perform one's civic duties." The sense of high political efficacy is associated with more exposure to political communication and with higher satisfaction with participation in the political process citizens (Almond and Verba, 1989: 198, 193). Since these classic studies, the concept of political efficacy has been refined. Niemi, Craig and Mattei (1991: 1407) distinguish between two different dimensions of the concept. Internal efficacy refers to "beliefs about one's own competence to understand, and to participate effectively in, politics" and external efficacy to "beliefs about the responsiveness of governmental authorities and institutions to citizen demands". These dimensions have been explored in many studies of political efficacy in the US (for an overview see Niemi, Craig and

Mattei, 1991: 1408), and in comparative studies (Hayes and Bean, 1993), and political efficacy questions are now standard items in national election surveys.

Empirical studies of the causes of individual citizens' sense of internal political efficacy show the importance of several individual-level factors, including gender, age, education and income (Almond and Verba, 1989; Hayes and Bean, 1993: 269-270). Jurisdiction size is also found to affect efficacy, but there is some variation in the empirical results. The sense of internal political efficacy is stronger towards local government than towards the national government (Almond and Verba, 1989: 141; Dahl and Tufte, 1973: 57). As Dahl and Tufte's theoretical arguments suggest, there are good reasons to expect similar differences between local units of varying size. Dahl and Tufte (op.cit.) report early evidence from Swedish local government which confirms the size-participation relationship, and this is confirmed by a recent investigation on Swedish data (Nielsen, 2003). In their analysis of the US, Almond and Verba (1963: 234) conclude that political efficacy does not depend on city size, but Finifter and Abramson (1975) argue that this conclusion is misleading; since education levels tend to be higher in larger cities, and since education is positively related to the sense of political efficacy, the true relationship between city size and political efficacy may be suppressed. When education is controlled for, it is revealed that city size is in fact negatively related to feelings of internal political efficacy (Finifter, 1970: 403-404; Finifter and Abramson, 1975: 194).

To sum up, many cross-sectional, and a few longitudinal, studies investigate the democratic consequences of differences in jurisdictions size, and there is some, but by no means unequivocal, evidence that larger size is associated with lower quality, however measured, of democracy. These studies are, however, subject to several potential methodological problems.

First, differences in observable, but unmeasured or badly measured, population characteristics can affect estimates, as in Finifter and Abramson (1975) discussed above. The populations of small and large jurisdictions can be different in many dimensions, leading to problems of common support rarely addressed in this literature. Education levels, mentioned above, are one important example, but also income levels, occupations and employment patterns

are different in small and large jurisdictions. It is hard to know whether differences in correlations are caused by jurisdiction size or simply by differences in the populations. This problem can, however, partly be solved by statistical control for the confounding variables.

Second, cross sectional studies cannot address problems of selection that arise if people with unobservable preferences over local democracy self-select into jurisdictions of different size, partly with an eye on exactly this. The fact that people may sort themselves into different jurisdictions is precisely the Tiebout idea (Ellickson, 1971; Ross and Yinger, 1999). While there is in general little empirical support to the hypothesis that people sort based on income (e.g. Rhode and Strumpf, 2003), solid evidence for sorting based on preferences for local public goods such as school quality and environmental quality is beginning to emerge (Banzhaf and Walsh, 2008). Such sorting is also found in the Danish case analyzed below (Kristensen, 2002). This constitutes a problem of endogeneity. When sorting effects exist, differences in political efficacy among citizens residing in jurisdictions of varying size may reflect reverse causality, that people with interests in local politics are drawn to smaller jurisdictions, rather than a causal effect of jurisdiction size on political efficacy.

The problem of endogeneity is difficult to handle statistically. One way to correct for endogeneity is to use an instrumental variable approach, but valid instruments are hard to find, and we have not seen any studies that attempt to use this approach to solve the problem. Another solution would be to study jurisdiction size and the quality of local democracy over time. This is, however, often infeasible, as important jurisdictional boundaries change only rarely. When change is observed in connection to large-scale reform, two additional complications arise: First, in large-scale reforms several factors often change at the same time. For example, in reforms aimed at increasing citizen empowerment, decentralization – an increase in the number of jurisdictions or devolution of powers to existing jurisdictions – can be accompanied by increasing transparency, changes in local political institutions and increases in funds from other levels of government or international institutions, making it difficult to isolate

the effects of changing jurisdiction size. Second, and related to the first, such reforms are often encompassing, leaving no individuals or jurisdictions unaffected.

In addition, when jurisdictional change happens, or is refrained from, it often does so endogenously. In many settings, jurisdictional change, whether at the city level (see Lassen, 2005), county level (see Alesina, Baqir and Hoxby, 2004) or country level (see Bolton and Roland, 1997, and Alesina and Spolaore, 1997) is often decided by voting. Alesina et al. (2004) finds that jurisdictions in the U.S. forego consolidation, by voting against it, if it means increasing population heterogeneity. Such majority, or supermajority, consent is necessary for consolidation in most countries and at different levels of government, and introduces the issue of selection bias into estimates of the effect of jurisdiction size on local democracy indicators.[7]

Therefore, in order to identify a causal effect of the size of the populace on political participation and political efficacy, it is necessary either to allocate people randomly across jurisdictions of different population size, which is something we do not do in this paper, or to identify a natural, or quasi, experiment which provides an exogenous source of variation in jurisdiction population size, which is what we do. In the next section we introduce the amalgamations of Danish municipalities, and argue that this provides an excellent opportunity for testing the effect of jurisdiction size on political efficacy. Before turning to the methodological considerations, we will introduce the Danish municipalities and give a short review of the existing Danish studies of the effects of jurisdiction size.

**II.1. The Danish System of Local Government**

---

[7] For a general criticism of using endogenously decided changes as providing variation for empirical identification, see Besley and Case (1999). A large literature in urban economics and politics in the US considers both case and large-N studies of various consolidation efforts (e.g. Rogers and Lipsey, 1974; Benton and Gamble, 1983; deHoog, Lowery and Lyons (1990)), but these efforts are prone to problems of this type of selection bias.

The local level is a vital part of the Danish public sector, both culturally and politically. Culturally, a majority of Danes first and foremost identify with their local area, rather than their region or country.[8] Politically, Danish local governments are responsible for about 40 per cent of Danish public expenditure, and relatively autonomous. They are free to set the rate of the local income tax and have wide discretion in most policy areas. Most notably, they control many important welfare services such as the school system, care for children and care for the elderly. The municipalities are governed by a city council elected in popular elections, in which turnout is generally high, for four year terms. The mayor is elected indirectly by the city council. Apart from presiding over the council and being the formal head of the municipality, the mayor is the only full-time politician, is head of the local administration, and is chairman of the economic committee which assumes a central role in the political process (see Berg and Kjær, 2005). The local party systems are typically dominated by local branches of the national parties, though with some so-called local lists.

Just as in Almond and Verba's results, the sense of political efficacy in Denmark tends to be higher towards the local level than towards the national level and the European Union, although the differences between the local and the national level are limited (Andersen, 2000: 128-130). In a thorough cross-sectional study of the effect of jurisdiction size on internal political efficacy in Danish municipalities, Lolle (2003) concludes that jurisdiction size does not have a strong effect on political efficacy. The association between characteristics of citizens and their sense of political efficacy is much stronger and, in particular, education is strongly correlated with internal political efficacy (Andersen, 2000: 141-142; Andersen, 2003: 116-118; Serritzlew, 2004).

---

[8] World Value Survey, 1999-wave. 55.8 per cent of respondents answered "Locality" to question of "Geographical groups belonging to first." Accessed February, 2009.

## III. The Danish Structural Reform

The Structural Reform (SR), implemented politically in 2005 and administratively in January 2007, dramatically consolidated Danish local government. The SR has three parts. First, the number of municipalities was reduced from 271 to 98,[9] increasing average jurisdiction size from 20,100 to 55,600 citizens. Second, 14 counties were merged into five new administrative units, so-called Regions. Third, new tasks were to be undertaken by the municipalities.

The SR was planned and implemented by the central government. After a short debate following the first announcement, predominantly in newspapers on the possibility of reform, the government in 2002 appointed the *Commission on Administrative Structure* with a chairman appointed by the Minister for the Interior, three independent expert members, four representatives from municipalities and counties, and four representatives from Danish ministries. No publicly elected officials participated in this work. The four volume report of the committee was published in January 2004. It recommended a large scale structural reform. Based on this, the centre-right minority government introduced a bill, which immediately got a majority when the support party Danish People's Party declared that they had one question for the negotiations with the government: Where shall we sign? After less than half a year, in June 2004, it was clear that Danish local government should experience dramatic reform.[10]

According to the agreement between the government and the Danish People's Party, the municipalities should be merged into larger units with at least 30,000 citizens. All municipalities with less than 30,000 citizens (230 of the 270 municipalities, see Table 1) were asked to arrange for mergers before January 1st 2005. Only neighbouring municipalities would be allowed to merge. Eleven of the new 98 municipalities had already before this agreement

---

[9] When the debate over the reform began in 2002, five smaller municipalities on the island of Bornholm voluntarily merged into one island-wide municipality, reducing the total number of municipalities from 275 to 271. These municipalities are not part of our sample.

[10] The reform process is documented in detail in Christiansen and Klitgaard's (2008) *The Unimaginable Reform* (in Danish: Den utænkelige reform).

announced how they wished to merge, and most of the remaining municipalities were very close to agreeing on with whom to merge (Blom-Hansen et al., 2006: 17-8). A few municipalities, especially islands, would be allowed to continue with less than 30,000 citizens, but these municipalities would have to cooperate with larger ones in providing some services to their citizens.


< Table 1 here >


Of the 270 original municipalities, 237 were merged into 65 new ones, and the remaining 33 municipalities continued without any changes. The first election involving the post-reform municipalities as political entities took place in November 2005, and the new structure took effect from January $1^{st}$ 2007. From this date the municipalities were made responsible for additional tasks (most importantly certain social services for handicapped, prevention, rehabilitation, and special education), and administration of tax collection was centralized. The transfer of tasks to the municipalities was limited, and amount to approximately DKR 25 bill compared to the total municipal expenditure of (in 2007) DKR 333 bill.


**III.1 The Structural Reform as a Natural Experiment**

Setting aside standard problems of cross-section analysis, the main problem of studying effects of jurisdiction size by comparing existing entities with varying population size is sorting effects, which are present if individuals' choice of municipality depends on something related to the dependent variable of interest. By exposing citizens to shocks of different sizes, since this was dependent on the possibilities of mergers given by neighbouring municipalities, the SR introduced an exogenous element into the size of the local polity experienced by citizens, and it is this exogeneity we use for identification of causal effects. Two potential sources of endogeneity remain, however.

13

First, individual citizens can respond to the new municipal structure by relocating. Individuals with strong preferences for political participation at the local level suddenly finding themselves in a large municipality may choose to move to one of the small municipalities. In practice, we do not consider this a major problem, in particular since the post-reform survey was conducted within a year after the reform, and we furthermore exclude recent movers from the analysis to focus on respondents with some accumulated experience of local politics specific to their current municipality.

Second, although individuals were not allowed to choose if and how their municipality should be merged, the municipalities and their citizens did have some, in some cases substantial, influence on the merger. For example, municipalities of citizens with a strong preference for local democracy may have aimed for a merger with a small municipality.[11] In practice, however, the municipalities were severely constrained by the size requirements and the condition that merging municipalities should be neighbours.

Some insight into the process of amalgamation can be obtained from the fact that referenda were held in 63 of the original 270 municipalities (Jørgensen, 2006: 168). Most of the referenda were concerned with adjustments of boundaries, but some involved the choice between a merger with one or another municipality. In our sample of 32 municipalities, referenda took place in six cases, marked with asterisks in Table A in the appendix. In these cases, the group of citizens did have a direct influence on what would happen to their municipality, but individual citizens, of course, only had a minuscule chance of affecting the result.[12]

Finally, we acknowledge that change, in itself, can alter feelings of internal political efficacy. The major components of the Structural Reform that did not involve changing

---

[11]  In our analysis, though, this would bias our analysis against finding a detrimental effect of the reform on local political efficacy.

[12] If we carry out our analysis below on the 26 municipalities without consultative referenda, overall results do not change substantively. We return to this in the results section.

jurisdictional boundaries are, as already discussed, controlled for by having a control group in the sample. However, for the amalgamated municipalities we can decompose the jurisdictional change into two components: The fact that jurisdictional boundaries changes at all and the fact that population size increases. If citizens react in a negative way to change per se, then any change in jurisdiction size, both positive and negative, should be associated with lower political efficacy. On the other hand, if population size is more important and has a systematic effect on political efficacy then increasing and decreasing population size should have differing effects, and the effect should be larger, the larger the change in population. While we will distinguish amalgamated municipalities by the change in size when carrying out the analysis below, we do not observe jurisdictions decreasing in size as a consequence of the Structural Reform. However, some evidence against the hypothesis that change in itself matters can be drawn from a project involving decentralization of the municipality of Copenhagen, the capital and by far the largest municipality of Denmark. In this case, survey respondents overwhelmingly favoured smaller units with respect to political efficacy, both internal and external, even if the proposal eventually failed due to concerns about economic efficiency (Klausen, 2001; Lassen, 2005).

## IV. Data

We use two individual level surveys, carried out by phone by a professional polling firm, on voter participation in local democracy. One was carried out before the SR, in 2001, the other was carried out in late 2007 and early 2008, a year after the SR was administratively implemented. A core set of identical questions were asked in both surveys, making it possible to compare pre- and post-SR political behaviour.

The 2001-survey was a sample of 4,700 residents in Denmark aged 18 or more. The overall response rate was 59 percent. An equal number of respondents were randomly selected in 60 of the then 275 Danish municipalities (Houlberg & Pedersen, 2003: 209). The 60

municipalities were randomly selected from six strata based on size, with ten municipalities from each.

The 2007-survey was a sample of 2,816 residents in Denmark. Of these, 709 either was younger than 24 years old, had non-functional phone numbers, or could not be reached in at least six attempts. Of 2107 contacted respondents, 59 percent completed the survey. [13]

Since our aim is to compare the two surveys, only residents aged 24 (who were 18 when the first survey was carried out) or more were included. Younger respondents were excluded in order rule out any cohort effects. Only municipalities selected in the 2001-survey were included in the 2007-survey. The 2007-survey was stratified by pre-SR municipalities, making it possible to compare municipal level responses in pre-SR municipalities within the same area, now part of a new, larger post-reform municipality. At the same time, we also surveyed municipalities unaffected by the reform.

The 60 municipalities of the 2001-survey were divided into five strata, one with continuing municipalities and four with reform municipalities. This latter group was constructed by combining below and above median size with below and above median population growth. We randomly selected eight municipalities from the control group stratum and six from each of the other four strata. Table 2 shows the strata and the 32 selected municipalities.


< Table 2 here >


The selected municipalities are listed in Table A in the appendix.

---

[13] See appendix for a detailed breakdown. While response rates are similar in the two samples, a general tendency to survey fatigue, observed both for Denmark and for other countries (Hansen, 2007) during the last decade, can distort patterns of response. However, our research design, by controlling for national level trends (in the case of matching even conditional on observables) using a control group addresses this problem fully.

We measure internal political efficacy by the five standard questions described in Niemi, Craig and Mattei (1991), combined into an additive index.[14] These are the questions used in the U.S. National Election Study. The questions has been translated into Danish and adapted to local politics. The (English translation) of the questions appear in Table 3.

< Table 3 here >

We use the exact same formulation in the 2007-survey as was used in the 2001-survey (see Lolle, 2003). Table 3 also shows descriptive statistics on unweighted samples and definitions of the independent variables used in the analyses below.

## V. Research design and empirical specification

The exogenous shock to jurisdiction size implied by the reform constitutes, together with the repeated survey data, the basis of our research design. To estimate the causal effect of jurisdiction size on internal political efficacy we use the differences-in-differences (DiD) method on repeated cross sectional data. DiD is well known in the program evaluation literature (see, e.g., Imbens and Wooldridge (2008) for a recent treatment).[15] Under DiD units or individuals receiving treatment are compared to their pre-treatment levels, while the same is done for a control group. This yields the first level of differences. Subsequently, these two differences are subtracted from each other, resulting in a second level difference, leaving an estimate of the causal effect of treatment under certain assumptions, which are detailed below.

### V.1. The Differences-in-differences estimator

[14] We use the index to facilitate comparison with the literature. Reliability is reasonably high, at .62. The empirical results generally hold the components *Qualified, Informed* and *Complex*, but not for *Understand* and *Public office*. In the matched sample, results are stronger and significant for all components but *Public office*. (Basic results are available for reviewers in appendix).

[15] In the early evaluation literature, this is also known as "interrupted time series with a nonequivalent no-treatment control group time series"-design. See Cook and Campbell (1979, p. 214).

Let the internal efficacy of an individual $i$ be $Y_i(1)$ in the treated case, that of municipal amalgamation, and $Y_i(0)$ be the outcome for the same person in case of no treatment. Our object of interest is the difference $\Delta_i = Y_i(1) - Y_i(0)$ which is the causal effect of municipal population size on the internal political efficacy experienced by individual $i$. In practice, however, an individual $i$ cannot be treated and non-treated at the same time, which means that either $Y_i(1)$ or $Y_i(0)$ will be missing. The difference $\Delta_i$ results from the pair of *potential outcomes*, a term coined by Rubin (1974), which is to be distinguished from the realized outcome, $Y_i$. Since a counterfactual for the realized outcome is not available at the individual level, the statistical approach is to estimate the missing variable from appropriate group means.

We study a repeated cross section model of individuals surveyed in 2001, well before the reform took place, and individuals surveyed in late 2007, two years after the first election of the newly formed municipalities and almost one year after the actual implementation. This model has the form

$$Y_i = \alpha + \gamma_M M_i + \gamma_P T_i + \tau_{DID} M_i T_i + \beta' X_i + u_i \quad \forall i \in I,$$

where $I$ is the set of respondents. $M_i \in \{0,1\}$ is an indicator for reform (as opposed to continuing) municipalities. Below, we also consider a treatment indicator of the form $M_i \in \{0, SS, SL, LS, LL\}$ where we distinguish treated municipalities both by their pre-reform size and their post-reform relative size; we return to this. $T_i \in \{0,1\}$ is a time period indicator equal to 0 before the reform and 1 after the reform. The interaction(s) $M_i T_i$ take(s) on the value 1 for reform municipalities after implementation, and zero otherwise. Finally, $u_i$ represents unobservable characteristics, and is assumed to be independent of treatment status conditional on observables and across periods; we return critically to this assumption below.

It can be useful to consider the conditional means for the four different groups: reform and non-reform municipalities, before and after reform. They are as follows:

18

$$
\begin{aligned}
\text{(A)}\ \ \bar{Y}_{00} &= E\left(Y \mid M=0, T=0, X\right) = \alpha + \beta'X \\
\text{(B)}\ \ \bar{Y}_{01} &= E\left(Y \mid M=0, T=1, X\right) = \alpha + \beta'X + \gamma_P \\
\text{(C)}\ \ \bar{Y}_{10} &= E\left(Y \mid M=1, T=0, X\right) = \alpha + \beta'X + \gamma_M \\
\text{(D)}\ \ \bar{Y}_{11} &= E\left(Y \mid M=1, T=1, X\right) = \alpha + \beta'X + \gamma_P + \gamma_M + \tau_{DID}
\end{aligned}
\tag{1}
$$

The Before-After estimator for the Treated, BAT, used when no control group is available, is given by (1.D) – (1.C):

$$
\bar{Y}_{11} - \bar{Y}_{10} = \gamma_P + \tau_{DID}
\tag{2}
$$

The BAT-estimator includes both the effect of treatment and the common trend or aggregate change. In our case, $\gamma_P$ includes the effects on internal political efficacy from overall trends and aspects of the reform common to all municipalities. To estimate consistently the effect of size, we need to subtract this common trend which is equal to the Before-After estimator of the Control group, BAC:

$$
\bar{Y}_{01} - \bar{Y}_{00} = \gamma_P.
\tag{3}
$$

The BAC estimates the effect of reallocation of government tasks as well as any aggregate time trends, possibly due to declining survey participation, in internal political efficacy with respect to local government; the latter may have increased or decreased in reaction to the substantial public debate on municipal politics and the Structural Reform. Similarly, increased fiscal pressure from central authorities on municipalities will also be captured by this trend, as long as it is similar across all municipalities.

Based on our set-up in (1), we get the difference-in-difference estimator as

$$
\begin{aligned}
\tau_{DID} &= \left(\bar{Y}_{11} - \bar{Y}_{10}\right) - \left(\bar{Y}_{01} - \bar{Y}_{00}\right) \\
&= \left(\bar{Y}_{11} - \bar{Y}_{01}\right) - \left(\bar{Y}_{10} - \bar{Y}_{00}\right)
\end{aligned}
\tag{4}
$$

where the first row is the BAT minus the BAC estimator, while the second row shows that we can also interpret the DiD-estimator as the difference between the two cross-sectional estimators.

The assumption that the trend in (2) is common for treatment and control groups alike is a key identifying assumption in the DiD-approach (see Blundell and Macurdy, 1999). We first proceed under the assumption that this assumption is satisfied; for example, as noted above, the reallocation of government tasks across vertical levels of government was the same for all municipalities. However, if the impact of the reform is heterogeneous with respect to observable characteristics, and the distributions of such observable characteristics differ across control and treatment groups, additional assumptions are needed to make sure a comparison is based on suitably similar groups; we return to this below, where we consider matching. The second key identifying assumption, denoted no composition bias (Blundell and Macurdy,1999: p. 1612), is that the populations considered are the same across time. At the population level, this means that no individuals change from treatment to control group and vice versa; this is sometimes problematic in studies of individual responses to changes in taxation or job training programs, but in our case this would be directly relevant only if people moved with the explicit aim of living in a treatment or control municipality, which seems highly unlikely. In addition, in our empirical analysis we exclude people in the post-reform sample that recently moved from another municipality. The assumption is more problematic when we consider the actual sampling of individuals; we return to this below.

When estimating individual level responses to aggregate level variables, here treatment status and other municipal-level variables, it is important to account for the possible covariance in errors. We do this by allowing for clustering at the (new) municipal level; this specification allows for even higher correlation in sub-clusters such as old municipal levels. Our number of sampled clusters is not sufficiently large for us to be able to refer to asymptotic results and we therefore report bootstrapped standard errors, resampled on clusters, and bias-corrected

confidence intervals (see Donald and Lang, 2007, and Cameron, Gelbach and Miller, 2008) in the tables below.[16]

## V.2. Non-random assignment: Differential surveys and differences in observables across treatment groups

In existing DiD-studies on repeated cross sectional data, the analysis is typically carried out as different waves of a particular representative survey, such as the GSS. In our case, we compare internal political efficacy measures across surveys carried out in the same municipalities, using exact similar worded questions, but with different sampling strategies. The existence of differential sampling probabilities across surveys raises the problem that the surveys are not stratified in the same way, and thus it is possible that controlling for confounding variables does not solve the problem of composition bias, the absence of which, as noted above, is necessary for identification of causal effects in a DiD-model. To solve this problem, we want to match respondents across surveys on a number of individual-level variables and their interactions.

At the same time, treatment and control municipalities may differ with respect to observable characteristics, which can lead to a violation of the common trend assumption if effects of the reform are heterogeneous across observables; as noted above, for example, an explicit aim of the SR was to create municipalities with a minimum of 30,000 inhabitants, which in itself affects the probability of treatment.[17] To solve this issue, we want to match survey respondents across treatment status. In sum, therefore, there are two assignments that may be non-random: assignment to treatment or control group and assignment to pre-reform or post-reform survey. To address this, we employ a variant of differences-in-differences with

---

[16] An alternative would be to use a hierarchical linear model, but this requires much stronger assumptions on the distributions of error terms relative to the OLS-method underlying the DiD approach and the matching procedures, as they are semi-parametric and non-parametric, respectively. See Primo et al. (2007) for a comparison of cluster-correction and HLM-approaches.

[17] This is an example of the criterion, stressed by Rubin (2008, p. 820), of including explicitly the reasons for treatment assignment at the level of the decision-maker.

matching (DiDM), an estimator developed in the program evaluation literature by Heckman et al. (1997, 1998). [18]

We proceed by propensity score matching.[19] While intuitively desirable, exact matching on a vector of characteristics is often infeasible due to the large number of possible combinations. As shown by Rosenbaum and Rubin (1983), however, if potential outcomes are independent of treatment conditional on *X*, which is a precondition for estimating a causal effect from the data anyway, then this is also the case conditional on the propensity score, defined as the probability of receiving treatment conditional on *X*.

As we have two assignments that are potentially non-random, we follow Blundell et al. (2004) in defining two propensity scores. Conditional on *X*, $P_M = \Pr\left(M = 1 \mid X\right)$ is the probability of being observed in a treated municipality and $P_T = \Pr\left(T = 1 \mid X\right)$ is the probability of being surveyed in the post-reform sample. Based on the propensity scores, we can now write the identifying common trend assumption as

$$E\left(Y_{it}^0 \mid P_M, P_T, M = 1, T = 1\right) - E\left(Y_{it}^0 \mid P_M, P_T, M = 1, T = 0\right)$$
$$= E\left(Y_{it}^0 \mid P_M, P_T, M = 0, T = 1\right) - E\left(Y_{it}^0 \mid P_M, P_T, M = 0, T = 0\right).$$

This allows the time effects to differ by *X* and assures that the distribution of observed characteristics is the same across all four samples.

In practice, we estimate the two propensity scores by running a probit regression of assignment status (treatment and time, respectively) on *X*. The resulting predicted values are the estimated propensity scores. Our point of departure is the group of individuals in the treated group, pre-treatment ($M = 1$, $T = 0$). To match these individuals with respondents in each of the

---

[18] Imbens and Wooldridge (2008) provide a recent overview of development on program evaluation, covering both matching and selection on unobservables, including the DiD-estimator.

[19] Dehejia and Wehba (2002) provide a practical introduction to propensity score matching and Imai (2005) provides an excellent introduction in a political science setting. Barabas (2004) and Lassen (2005) use it in quasi-experimental settings to evaluate effects of deliberation on policy opinions and effects of information on turnout, respectively.

three control groups, we perform a variant of nearest neighbor caliper matching with

replacement by pairing each treated respondent with a respondent in each of three control

groups. The respondent in each control group is chosen to minimize the distance, using

Mahalanobi's metric, from the treated individual's two estimated propensity scores, given

restrictions on the maximum allowable distance (the caliper) both for the two-dimensional

distance measure and for the two one-dimensional propensity scores as well as a common

support restriction.[20] Based on these matched samples, we carry out standard difference-in-

difference analysis to form estimates of the Average Treatment effect on the Treated (ATT)

under the assumption of separable additivity of time and group effects.[21] Since matching is

carried out with replacement, we correct the sample sizes to reflect the weights implied by this

procedure.

Throughout, we, as is standard in the literature, report bootstrapped standard errors for

the estimated ATTs based on the DiDM procedure. A recent study shows that this may in fact

result in biased estimates for the standard error, but the issue is not resolved (Abadie and

Imbens, 2007). In performing the bootstrap, we resample at the cluster level of new

municipalities (see above).

### V.3. Unobserved heterogeneity in treatment

While control group municipalities generally preserved the status quo, the actual course of the

reform process in the treatment municipalities can potentially vary considerably, possibly as a

consequence of unobserved characteristics of the pre-reform municipality. To account for this,

we allow in the matching analysis for exact matching within the treatment group (across time)

---

[20] Blundell et al. (2004) identifies the nearest neighbor using the Euclidean distance. Mahalanobi's metric
takes into account the covariance between the propensity scores and results in slightly fewer matched
observations, but in practice the difference in results is negligible. Results are available in appendix for
reviewers.

[21] This is reminiscent of, but not identical to, the pre-processing of data suggested by Ho et al. (2007). As
they note (p. 205) their procedure is not directly applicable in the case of multiple treatments.

based on the pre-reform municipalities. This means that the matching procedure is carried out under the restriction that an observation in the pre-reform treatment group must be found in the same pre-reform municipality, rather than in the post-reform treatment group as a whole.[22]

As emphasized by Heckman et al. (1997) and Smith and Todd (2005) in their work on causal effects of job training programs, and confirmed more generally by Cook, Shadish and Wong (2008), DiDM functions best when matching individuals who reside within a given geographical unit and, if possible, are administered identical surveys. This is exactly what we do here.[23]

## V.4. Retrospective evaluations

To supplement the evidence from the repeated cross sections, we also asked survey respondents in the 2007 survey to assess, in retrospect, the consequences of the Structural Reform for measures of local political efficacy. While there are well-known issues with retrospective questions of this kind,[24] they allow us to get at the missing panel data in a different way: by asking one person to evaluate the differences before and after the reform, we essentially control for individual level factors and get a direct estimate of the first level differences.

We can now compare these differences across treatment and control municipalities as well as within the group of treatment municipalities. We analyze these data using both a regression-based cross-sectional framework and matching across treatment and control groups, for both the full sample and the matched sample resulting from the DiDM-

---

[22] An alternative would be to treat each of the municipal treatment categories, or even the municipalities themselves as separate treatments in the DiDM-analysis, but sample size makes this infeasible.

[23] Morton and Williams (2008, chapter 10) provide an introduction to the structural approach to causality and its focus on unobserved heterogeneity in a political science context. As noted above, the DiD-estimator is essentially structural, as noted by Blundell and Macurdy, even if most applications of it in a natural experiment context present it as a non-structural approach.

[24] These questions were purposely asked *after* the standard efficacy level questions, so as not to prime respondents.

analysis. Throughout we take into account the clustered nature of the data by allowing for within-municipality dependence in estimated standard errors.

## VI. Results

### VI.1. Regression-based DiD-estimates

Table 4 reports results from the standard, regression-based difference-in-difference analysis, for three different specifications: analysis of binary, categorical and continuous treatments on the full samples.

< Table 4 here >

The results for the binary treatment, where citizens in all reform municipalities are considered identically treated, suggest strong effects of the reform. First, the estimate on the post-reform variable suggests a general significant and negative effect of time on all municipalities, owing possibly to the overall reform or to changes in survey response pools; we return to this below. Second, there seems to be no significant differences across treatment and control municipalities before the reform. Turning to the ATT, the effect of the reform on citizen political efficacy in treated municipalities, we estimate the difference-in-difference effect to be -.77 for the full sample, bringing the total effect of reform for the treated municipalities to approximately -1.3. The DiD-estimate is significant at the 5%-level.[25]

Column 2 presents results from the categorical treatment. In this case, we distinguish municipalities by the pre-reform size and their post-reform relative size, creating the four categories SS, SL, LS and LL presented in table 2 above. This way of looking at the data

---

[25] As noted earlier, all standard errors are corrected for clustering at the new municipal level. Since the number of clusters is not large, reported standard errors are the results of a bootstrap procedure with cluster-level resampling and 200 replications.

shows quite different results for the four treatment categories: Respondents living in pre-reform small municipalities that post-reform are a relatively small share of total new municipality size experience the largest decline in internal political efficacy, a decline which is significant at the 1 % level, while respondents in the three other categories experience almost equivalent declines in internal political efficacy.

Finally, columns 3 and 4 report results from the continuous treatment, where we consider the effect of the change in population size experienced by respondents. Column 3 reports results when using the absolute change in population size, while column 4 reports results when using the relative change. Population size difference corresponds in a monotonic way to the categorical treatment variables used in column 2 and, indeed, we find results similar to those reported for the categorical case: The loss of internal political efficacy experienced by respondents is increasing in the population increase, both absolute and relative, following municipal amalgamation.

In addition to statistical significance, the results also have substantive significance, measured relative to other variables important for explaining internal political efficacy. While the average effect reported in column 1 is not large, as it includes responses from citizens in large municipalities that experienced very little change, the estimate reported for citizens from small municipalities that, post-reform, constituted a small share of the new municipality (equal to -1.1) is equivalent to the sample differences between men and women (also equal to -1.1) or the difference between having no education beyond high school and a college degree (also equal to 1.1).

Finally, we noted above that some municipalities held consultative referenda as part of the reform process. This is potentially important as the very act of a referendum may have increased the voters' feeling of legitimacy of the final amalgamation decision, which could affect their sense of internal political efficacy. Moreover, the decision to have a referendum in the first place could reflect a municipality more interested in local politics. Six municipalities in

our sample held referenda, but regression results for the non-referendum sub-sample are not affected by the exclusion from the analysis of the municipalities which held referenda.[26]

**VI.2. Results from differences-in-differences with matching**

Figure 1 presents scatter plots of the two propensity scores. Each observation in the figure is an individual, and we observe a wide range of observed propensity scores in both dimensions, documenting substantial differences in the probability of being observed in samples both across time and across treatment status.[27]

< Figure 1 here >

As described above, each observation in the treatment, pre-reform group is matched to its nearest neighbor measured by Mahalanobi's distance in each of the three control groups, subject to a maximum distance, and the samples are trimmed to the common support. This procedure results in 1253 observations being dropped from the sample, leaving 739 of which some will be weighted several times as sampling in the nearest neighbor matching was being done with replacement. [28] In standard one-dimensional propensity score analysis, it is common to focus on the post-matching densities of propensity scores and the resulting covariate balance between treatment and control groups, testing for significant differences in the latter. However, this is less straightforward in the present case, with four comparison groups and two propensity scores.

---

[26] Sample size precludes matching analysis in this context. On average, referenda municipalities report higher political efficacy than non-referenda municipalities, driven primarily by the municipality of Aalborg, which pre-reform was the fourth largest metropolitan area in Denmark. Results are available upon request (in appendix for reviewers).

[27] Individuals with almost zero probability of being observed in the reform municipalities are individuals living in the municipality of Copenhagen which differs in size from most of the other municipalities in the sample by a factor 20.

[28] In practice, we choose the maximum allowable distance, the caliper $c$, for the Mahalanobian distance to be .025, but results are not sensitive to this choice. With $c = .05$ we retain 920 observations, and with $c = .1$ we retain 1091 observations. In the latter case, results begin approaching the estimates from the full sample specification. Results are available upon request (in appendix for reviewers).

Visual inspection of covariate means across groups suggests that, overall, matching considerably improves covariate balance, but in a few cases balance seems to deteriorate. To give some sense of the differences, consider figure 2.

< Figure 2 here >

In figure 2, we present covariate balance before and after matching using a radar plot. In a perfectly balanced sample, the plotted figure is a (tilted) square; the further away from a square is the quadrangle, the worse is the balance. Consider the bottom graph on the left: The blue line says that mean populations in the control group samples before and after reform (groups 1 and 2) are 90.000 and 80.000, respectively, while they are approximately 25.000 in the treatment groups. Post-matching covariate means, shown by the solid, red line, shows an almost perfect square, with mean population for each of the four groups around 20.000.

Table 5 shows results from the matching analysis for four different specifications: Binary treatment, categorical treatment, population change and relative population change. The top panel reports results from the specification where there is no restriction on matching within the treatment group across time. The binary ATT is -1.0 which is significant at the 5 percent level and a 40 percent increase relative to the estimate from the full-sample regression-based result reported in table 4. As above, the categorical treatment regime suggests large differences across categories: The largest, and most precisely estimated, effect is for individuals coming from small pre-reform municipalities that constitute a small share of the post-reform municipality. The estimated ATT for these individuals is -1.8. The pattern generally is the same as in the regression-based analysis, with small and insignificant effects for citizens living in municipalities where there was little change ex post, as their pre-reform municipality remained the larger one.

< Table 5 here >

A similar picture emerges from the continuous measures: estimates of the effect of population difference and relative population difference increase by around 60 percent compared to the full sample case.

In the lower part of the table, we show results when the matching of treated individuals across time was done under the additional restriction that a match should be found in the same pre-reform municipality. This means that a survey respondent in the post-reform treatment group living in a particular area of a larger amalgamated municipality should be matched to a respondent living in the same area, which pre-reform was a municipality of its own. This restriction causes then total number of observations across the four cells to decrease by 25 percent, as it is now impossible to find suitable matches for some pre-reform treated respondents. The difference in results, however, is not large. The estimate of the binary ATT increases to -1.3, driven by a large increase in the estimate of ATT_SL.

The overall conclusion is clear: The qualitative results of the regression-based DiD-analysis remain valid, but taking into account non-random assignment across samples, in both the treatment and the time dimensions, increases the estimated effects of municipal amalgamation considerably. The estimated ATT from the weighted sample with exact matching on pre-reform municipalities (equal to -1.3) is an increase of more than 70 percent over the regression-based approach from table 4 (equal to -0.8).

Furthermore, in results not shown (available in appendix for reviewers), we find, in accordance with the full sample results, a significant overall decrease in political efficacy as a consequence of the reform of between .6 and .8 depending on specification, resulting in an estimated average total effect of the entire reform on the amalgamated municipalities of around -2.0.

## VI. 3. Retrospective evaluations

Table 6 shows the results from the retrospective questions. Each part of the table shows the result of an estimation process, either in the form of linear regression or, for the bottom row,

29

propensity score kernel matching. All estimations, including estimation of the propensity score, include a full set of controls as above.

< Table 6 here >

For all estimation methods and samples we consistently find a statistically significant decline in perceived political efficacy for individuals in reform municipalities. The results on the binary treatment indicator, reported in the top row, suggest reasonably similar effects across samples, equivalent to one half standard deviation. Identical results are obtained using kernel matching, reported in the bottom row.

The effect is not homogenous across treatment categories, however: The impact of reform is larger for individuals residing in small pre-reform municipalities and generally small and insignificant for inhabitants of the largest reform jurisdictions, when these jurisdictions constitute a large part of the newly formed municipality. Throughout, *F*-tests reject the null hypothesis of equal parameter estimates for the four categories. These results are confirmed by the continuous treatment variables: The effect of amalgamation is larger, the larger is the change in population size resulting from the reform; thus, individuals experiencing larger (both absolute and relative) changes in the size of their jurisdiction reported greater losses of perceived political efficacy.

The retrospective evaluations also allow us assess whether other factors associated with the amalgamation process influence individual evaluations of the change in political efficacy. If, for example, amalgamation changed local public finances dramatically, individual evaluations of the change in political efficacy could be affected by (dis)satisfaction related to this fact. We construct two variables to account for economic and political change: Economic change is measured by the difference in municipal tax base per capita as a consequence of the reform.[29] The change in tax base reflects both the amalgamation between

---

[29] Municipal tax base per capita is the standard metric for comparing local public finances and is a key variable in the Danish system of intergovernmental grants.

municipalities with different pre-reform tax bases as well as the change induced by the concurrent adjustment of the system of redistributive intergovernmental grants. Political change is measured by an indicator variable capturing whether a voter reporting to vote for the mayor's party pre-reform experienced a loss of political control for that party post-reform.

Economic change was very limited, partly as a result of the change in intergovernmental grants. On average, tax bases per capita in our sample increased by only 1.1 percent, with no discernable difference between control and treatment municipalities. This is also reflected in the results (not reported, available in appendix for reviewers) from including tax change, both absolute and relative, as determinants of the change in political efficacy. Tax base change is insignificant and does not affect results on population change. Similarly, individuals seeing their favourite party losing the mayor's office do not report significantly different levels of efficacy, and results on population change are likewise not affected. In sum, the reform did not cause large changes in the economic environment or political control of municipalities, and the changes did not affect neither levels of political efficacy on their own, nor the relationship between population change and political efficacy.

## VII CONCLUSION

Jurisdiction size affects citizens' sense of possibilities of participating in local politics in a causal fashion. Our estimates of the effects of jurisdiction size on internal political efficacy are based on survey data collected before and after a recent comprehensive administrative reform of Danish local government, and we find the effects to be substantial: for individuals residing in small jurisdictions pre-reform, increases in municipal size reduced levels of internal political efficacy by a magnitude similar to existing differences between men and women or high-school and college graduates.

Our identification strategy is based on two key features of the reform: First, the reform was largely exogenous from the viewpoint of individual citizens and municipalities,

31

which enables us to address problems of sorting and endogenous political change not considered in the existing literature. Second, it affected only some jurisdictions, allowing us to control, using a difference-in-difference approach, for other concurrent changes. The results are stable across different estimators and samples, including two-dimensional propensity score matching to address directly the structural assumptions underlying the DiD-estimator, and across different measures of political competence.

How does jurisdiction size affect feelings of internal political efficacy? In the literature, explanations have focused on both individual resources such as education in dealing with costs of participation, including information processing, and on differences in psychological engagement and mobilization patterns. In his study of participation in metropolitan America, Oliver (2000) addresses exactly this issue: he does not find support for the hypothesis that costs of participation, modelled as interactions between city size and individual resource variables, can explain the negative relationship between size and civic involvement identified in his analysis. However, he does find significant differences in levels of interest in local politics and patterns of political mobilization. Obviously, picking up on the discussion above, this may reflect sorting of people with different levels of political interest into jurisdictions of political size rather than a causal effect of size on political interest.

The matching analysis carried out here provides a natural first step in estimating differences in responses to changes in municipal size. In on-going work (Lassen and Serritzlew, 2009), we observe substantial treatment heterogeneity across some individual factors, emphasizing again that the estimates reported above are *average* treatment effects. Our preliminary analysis suggests no effect of education, but reveals large and significant effects of population size change for men's feeling of internal political efficacy, while we see no effect for women. In a similar way, there seems to be little effect of the reform on the political efficacy for the young and the old. One potential reason is that on average women, as well as the young and the old, pay less attention to local politics in the first place and, therefore, are less likely to

observe both positive and negative changes in the political setting. As a result, political consolidation may in fact reduce, rather than enhance, inequality in political participation.

Finally, the results of this paper have implications for research into the architecture of government: First, the results contribute to the on-going debate about the optimal level of decentralization (Treisman, 2007), by substantiating the causal argument of Dahl and Tufte that having smaller jurisdictions improves local democracy in a causal fashion. Second, we believe our approach, explicitly addressing issues of sorting and endogeneity identified in the literature on urban economic geography (Tiebout, 1956; Ross and Yinger, 1999; Alesina et al. 2004), can be applied to other parts of the (de)centralization trade-off outlined in the introduction, specifically other dimensions of local democracy, service delivery and economies of scale, providing a micro-level foundation for decisions on decentralization in developed and developing economies alike.

**REFERENCES**

Abadie, A. and G. Imbens. 2007. "On the Failure of the Bootstrap for Matching Estimators." *Econometrica*, forthcoming.

Alesina, A. and E. Spolaore. 1997. "On the Number and Size of Nations." *Quarterly Journal of Economics* 112, 1027-1056.

Alesina, A. and E. Spolaore. 2003. *The Size of Nations*. Cambridge, MA: MIT Press.

Alesina, A., R. Baqir and C. Hoxby. 2004. "Political Jurisdictions in Heterogeneous Communities." *Journal of Political Economy* 112, 348-396.

Almond, G. A. and S. Verba. 1963/1989. *The Civic Culture.* Princeton, N.J.: Princeton University Press

Andersen, J. G. 2000. "Magt og afmagt: Nyt perspektiv på 'political efficacy'", pp. 123-152 in Jørgen Goul Andersen, Lars Torpe and Johannes Andersen (eds.) *Hvad folket magter. Demokrati, magt og afmagt*, Købehavn: Jurist- og Økonomforbundets Forlag

Andersen, J. G. 2003. *Over-Danmark og under-Danmark? Ulighed, velfærdsstat og politisk medborgerskab*. Århus: Aarhus Universitetsforlag.

Banzhaf, H. S. and R. P. Walsh. 2008. "Do People Vote with Their Feet? An Empirical Test of Tiebout's Mechanism." *American Economic Review* 98, 843-863.

Barabas, J. 2004. "How Deliberation Affects Policy Opinions**.**" *American Political Science Review* 98, 687-701.

Barankay, I. and B. Lockwood. 2007. "Decentralization and the productive efficiency of government: Evidence from Swiss cantons." *Journal of Public Economics* 91, 1197-1218.

Benton, J.E. and D. Gamble. 1983. "City/County Consolidation and Economies of Scale: Evidence from a Time-Series Analysis in Jacksonville, Florida." *Social Science Quarterly* 65, 190-198.

Berg, R. and U. Kjær. 2005. *Den danske borgmester*, Odense: Syddansk Universitetsforlag.

Besley, T. and S. Coate. 2003. "Central versus local provision of public goods: a political economy analysis." *Journal of Public Economics* 87, 2611–2637.

Blom-Hansen, J., J. Elklit & S. Serritzlew. 2006. "Den store kommunalreform og dens konsekvenser", pp. 11-36 in J. Blom-Hansen, J. Elklit & S. Serritzlew (eds.), *Kommunalreformens konsekvenser*, Århus: Academica.

Blundell, R. and T. Macurdy. 1999. "Labor Supply: A Review of Alternative Approaches." Chapter 27 in *Handbook of Labor Economics* Volume 3A, 1559-1695.

Blundell, R., M. Costa Dias, C. Meghir and J. Van Reenen. 2004. "Evaluating the Employment Impact of a Mandatory Job Search Program." *JEEA* 2, 569-606.

Bolton, P. and G. Roland, 1997. "The Breakup of Nations: A Political Economy Analysis." *Quarterly Journal of Economics* 112, 1057-1090.

Bordignon, M. F. Cerniglia and R. Revelli. 2004. "Yardstick competition in intergovernmental relationships: theory and empirical predictions." *Economics Letters* 83, 325–333.

Brady, H. E. and J. McNulty. 2004. "The Costs of Voting: Evidence from a Natural Experiment." Presented to the 2004 Annual Meeting of the Society for Political Methodology, Stanford.

Cameron, A.C., J. Gelbach and D. Miller. 2008. "Bootstrap-based improvements for inference with clustered errors." *Review of Economics and Statistics* 90, 414-427.

Campbell, A., G. Gurin and W. E. Miller. 1954. *The Voter Decides.* Evanston: Row, Peterson.

Campbell, D.T. 1969. "Reforms as Experiments." *American Psychologist* 24, 409-29.

Christiansen, P.M. and M.B. Klitgaard. 2008. *Den utænkelige reform.* Odense: Syddansk Universitetsforlag. (*The Unthinkable Reform*)

Cook, T. and D. Campbell. 1978. *Quasi-Experimentation.* Boston: Houghton Mifflin.

Cook, T., W. Shadish and V. Wong. 2008. "Three Conditions under Which Experiments and Observational Studies Produce Comparable Causal Estimates: New Findings from Within-Study Comparisons." *JPAM* 27, 724-750.

Craig, S. C., R. G. Niemi and G. E. Silver. 1990. "Political Efficacy and Trust: A Report on the NES Pilot Study Items", *Political Behavior*, Vol. 12, No. 3, pp. 289-314.

Dahl, R. A. and E. R. Tufte. 1973. *Size and Democracy.* Stanford University Press.

Dehejia, R. and S. Wahba. 2002. "Propensity Score Matching Methods for Nonexperimental Causal Studies." *Review of Economics and Statistics* 84, 151-161.

DeHoog, R., D. Lowery and W. Lyons. 1990. "Citizen Satisfaction with Local Governance: A Test of Individual, Jurisdictional, and City-Specific Explanations." *Journal of Politics* 52, 807-837.

Donald, S. and K. Lang. 2007. "Inference with differences-indifferences and other panel data." *Review of Economics and Statistics* 89, 221-233.

Ellickson, B. 1971. "Jurisdictional Fragmentation and Residential Choice." *American Economic Review* 61, 334-9.

Faguet, J-P. 2004. "Does Decentralization Increase Responsiveness to Local Needs? Evidence from Bolivia." *Journal of Public Economics* 88, 867-894.

Finifter, A. W. and P. R. Abramson. 1975. "City size and Feelings of Political Competence", *Public Opinion Quarterly*, vol. 39, no. 2, pp. 189-198

Finifter, A. W. 1970. "Dimensions of Political Alienation", *American Political Science Review*, vol. 64, no. 2, pp. 389-410.

Hansen, Kasper M. 2007. "The effects of incentives, interview length, and interviewer characteristics on response rates in a CATI-study." *International Journal Public Opinion Research* 19, 112-21.

Hayes, B. C. and C. S. Bean. 1993. "Political efficacy - a comparative-study of the United States, West Germany, Great Britain and Australia", *European Journal of Political Research*, vol. 23, pp. 261-280.

Heckman, J., H. Ichimura, J. Smith, P. Todd. 1998. "Characterizing Selection Bias Using Experimental Data." *Econometrica* 66, 1017-1098.

Heckman, J., H. Ichimura, P. Todd. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." *Review of Economic Studies* 64, 605-654.

Hindriks, J. and B. Lockwood. 2005. "Decentralization and electoral accountability: incentives, separation and voter welfare." *CEPR Discussion Paper 5125*.

Hooghe, L. and G. Marks. 2003. "Unraveling the Central State, but How? Types of Multi-Level Governance." *American Political Science Review* 97, 233-243.

Ho, D. E., K. Imai, G. King and E. A. Stuart. 2007. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 15, 199–236.

Houlberg, K. & C. S. Pedersen. 2003. "Appendiks 1: Dataindsamling", pp. 207-210 in Ulrik Kjær & Poul Erik Mouritzen (eds.), *Kommunestørrelse og demokrati*, Odense: Syddansk Universitetsforlag.

Imai, K. 2005 "Do Get-Out-The-Vote Calls Reduce Turnout? The Importance of Statistical Methods for Field Experiments." *American Political Science Review* 99, 283-300.

Imbens, G. and J. Wooldridge. 2008. "Recent Developments in the Econometrics of Program Evaluation." NBER Working Paper #14251, August.

Jørgensen, L. P. 2003. "Hvor skal vi hen? Folkeafstemninger: Eliternes instrument eller folkets stemme?", pp. 167-187 in Poul Erik Mouritzen (ed.), *Stort er godt*, Odense: Syddansk Universitetsforlag.

Kjær, U. and P. E. Mouritzen. 2003. "Er der sammenhæng mellem kommunestørrelse og lokalt demokrati?", in Kjær and Mouritzen (eds.) *Kommunestørrelse og lokalt demokrati*, Odense: Syddansk Universitetsforlag.

Klausen, J. E. 2001. *Bydelsråd i København: En analyse av folkeavstemningen.* Oslo: NIBR Prosjektrapport 2001:2.

Kristensen, T. B. 2002. "Kommunestrukturens konsekvenser - politik, økonomi eller politisk økonomi?", in Blom-Hansen, J., Bruun, F., Pallesen, T. (eds.) *Kommunale patologier*, Systime, Århus.

Lassen, D. D. 2005. "The Effect of Information on Voter Turnout: Evidence from a Natural Experiment." *American Journal of Political Science* 49, 103-18.

Lolle, H. 2003. "Kommunalpolitisk selvtillid", pp. 162-176 in Ulrik Kjær and Poul Erik Mouritzen (eds.), *Kommunestørrelse og lokalt demokrati*, Odense: Syddansk Universitetsforlag.

Lyons, W. and D. Lowery. 1989. "Governmental Fragmentation versus Consolidation: Five Public-Choice Myths about How to Create Informed, Involved, and Happy Citizens." *Public Administration Review* 49, 533-543.

Meguid, B. 2007. "Bringing Government Back to the People? The Impact of Political Decentralization on Voter Engagement in Western Europe." Manuscript, presented at 19th Biennial Meeting of the European Union Studies Association, Montreal, Canada, May.

Morlan, R. 1984. "Municipal vs. National Election Voter Turnout: Europe and the United States." *Political Science Quarterly* 99, 457-470.

Morton, R. and K. Williams. 2008. *From Nature to the Lab: Experimental Political Science and the Study of Causality.* Book manuscript, NYU, September 15.

Niemi, Richard G., Stephen C. Craig and Franco Mattei. 1991. "Measuring Internal Efficacy in the 1988 National Election Study", *American Political Science Review*, vol. 85, no. 4, pp. 1407-1413.

Nielsen, P. 2003. *Kommunindeling och demokrati. Om sammenlaggning og delning av kommuner i Sverige*. (Municipality division and democracy. Amalgamation and secession of municipalities in Sweden). Acta Universitatis Upsaliensis, *Skrifter utgivna av Statsvetanskapliga foreningen i Uppsala* no. 157.

Oates, W.E. 1972. *Fiscal Federalism*. NY: Harcourt Brace Jovanovich.

Oliver, J. E. 2001. *Democracy in Suburbia*, Princeton, NJ: Princeton University Press.

Oliver, J. E. 2000. "City Size and Civic Involvement in Metropolitan America." *American Political Science Review* 94:361-73.

Primo, D. M., M. L. Jacobsmeier and J. Milyo. 2007. "Estimating the Impact of State Policies and Institutions with Mixed-Level Data." *State Politics and Policy Quarterly* 7, 446-59.

Rhode, P.W. and K. Strumpf. 2003. "Assessing the Importance of Tiebout Sorting: Local Heterogeneity from 1850 to 1990." *American Economic Review* 93, 1648-1677.

Rogers, B.D. and C.M. Lipsey. 1974. "Metropolitan Reform: Citizen Evaluations of Performances in Nashville-Davidson County, Tennessee." *Publius* 4, 19-34.

Rosenbaum, P. and D. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrica* 70, 41-55.

Ross, S. and J. Yinger. 1999. "Sorting and Voting: A Review of the Literature on Urban Public Finance." Chapter 47 in *Handbook of Regional and Urban Economics*, Amsterdam: Elsevier.

Rubin, D. 1974. "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of Educational Psychology 66*, 688-701.

Rubin, D. 2008. "For Objective Causal Inference, Design Trumps Analysis." *Annals of Applied Statistics* 2, 808-840.

Seabright, P. 1996. "Accountability and decentralization in government: an incomplete contracts model." *European Economic Review* 40, 61–91.

Serritzlew, S.. 2004. "Det lokale demokrati i den nye kommunestruktur", *Fokus-Nyt*, no. 27, pp. 4-9.

Smith, J. and P. Todd. 2005. "Does matching overcome LaLonde's critique of nonexperimental estimators?" *Journal of Econometrics* 125, 305-353.

Stenner-Day, K. and M. Fischle (1992), "The Effects of Political Participation on Political Efficacy: A Simultaneous Equations Model", *Australian Journal of Political Science*, vol. 27, pp. 282-305.

Tiebout, C. 1956. "A Pure Theory of Local Expenditures." *Journal of Political Economy* 64, 416-24.

Treisman, D. 2007. *The Architecture of Government: Rethinking Political Decentralization.* Cambridge University Press.

Verba and Nye, 1972.

World Bank. 2000. "Entering the 21st Century" in *World Development Report 2000*, Washington D.C.

**Table 1: Danish municipalities before and after the Structural Reform**

| Number of citizens | Municipalities pre-reform (2006) | | Municipalities post-reform (2007) | |
|---|---|---|---|---|
| | Frequency | Percent | Frequency | Percent |
| 0-5,000 | 14 | 5 | 3 | 3 |
| 5,001-10,000 | 114 | 42 | 1 | 1 |
| 10,001-20,000 | 77 | 29 | 3 | 3 |
| 20,001-30,000 | 25 | 9 | 18 | 18 |
| 30,001-50,000 | 23 | 9 | 39 | 40 |
| 50,001-100,000 | 13 | 5 | 28 | 29 |
| More than 100,000 | 4 | 1 | 6 | 6 |
| Total | 270 | 100 | 98 | 100 |

Source: Statistics Denmark.

**Table 2: Strata in the post-SR survey**

| Stratum | Merged? | Pre-SR size | Population growth | No. of municipalities | Share of sample | Share complete answers |
|---|---|---|---|---|---|---|
| 1 (non merged) | No | N/A | N/A | 8 | 25,0 % | 25,1 % |
| 2 (Small – small part) | Yes | Below median | Above median | 6 | 18,7 % | 18,7 % |
| 3 (Small – large part) | Yes | Below median | Below median | 6 | 18,8 % | 18,7 % |
| 4 (Large – small part) | Yes | Above median | Above median | 6 | 18,8 % | 18,7 % |
| 5 (Large – large part) | Yes | Above median | Below median | 6 | 18,8 % | 18,8 % |
| Total | | | | 32 | 100,0 % | 100,0 % |

**Table 3: Questions, definitions, and descriptive statistics**

| Variable | Question or definition | Pre-SR survey | | | | | Post-SR survey | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Independent variables | | N | Mean | Std. dev. | Min | Max | N | Mean | Std. dev. | Min | Max |
| Vocational training | Respondent has completed vocational training | 2763 | .25 | .43 | 0 | 1 | 1233 | .34 | .48 | 0 | 1 |
| College degree | Respondent has completed two year theoretical education | 2763 | .19 | .39 | 0 | 1 | 1233 | .26 | .44 | 0 | 1 |
| Master's degree or above | Respondent has received Master's degree or higher | 2763 | .08 | .27 | 0 | 1 | 1233 | .13 | .34 | 0 | 1 |
| Income | Pre-tax income of household in categories | 2281 | 2.7 | 1.2 | 1 | 5 | 1233 | 2.7 | 1.3 | 1 | 5 |
| Publicly employed | Employed in the public sector | 2763 | .36 | .48 | 0 | 1 | 1233 | .25 | .43 | 0 | 1 |
| Gender | Gender of respondent (1= female) | 2763 | .52 | .50 | 0 | 1 | 1233 | .50 | .50 | 0 | 1 |
| Age | Age of respondent at time of survey | 2761 | 48 | 16 | 17 | 96 | 1233 | 53 | 15 | 24 | 95 |
| Dependent variable: Internal political efficacy | Additive index of the following five variables | 2450 | 16 | 4.6 | 5 | 25 | 1174 | 16 | 3.9 | 5 | 25 |
| Qualified | I consider myself to be well qualified -to participate in municipal politics | 2716 | 2.8 | 1.6 | 1 | 5 | 1225 | 2.8 | 1.4 | 1 | 5 |
| Understand | I feel that I have a pretty good under-standing of the important political issues facing our municipality | 2612 | 3.8 | 1.2 | 1 | 5 | 1212 | 3.8 | 1.1 | 1 | 5 |
| Public office | I feel that I could do as good a job as a local councilor as most other people | 2660 | 2.9 | 1.6 | 1 | 5 | 1208 | 2.9 | 1.4 | 1 | 5 |
| Informed | How well would you say that you are informed about local politics in your municipality? | 2728 | 3.4 | .93 | 1 | 5 | 1233 | 3.2 | 1.0 | 1 | 5 |
| Complex | Sometimes local politics seem so complicated that a person like me can't really understand what's going on | 2679 | 3.4 | 1.5 | 1 | 5 | 1213 | 3.2 | 1.3 | 1 | 5 |

Note: The table reports summary statistics from the raw sample, without weighting.

**Table 4: Regression-based DiD-analysis of the effect of municipal amalmagation and size on political efficacy**

| | Dependent variable: Internal political efficacy | | | |
|---|---|---|---|---|
| Postreform (P = 1) | -0.555 | -0.557 | -0.601 | -0.782 |
| | [0.258]** | [0.258]** | [0.257]** | [0.208]** |
| Treated municipalities (M = 1) | 0.093 | | | |
| | [0.292] | | | |
| Treated Ms, post-reform (P x M = 1) | -0.772 | | | |
| | [0.281]** | | | |
| Treated municipalities (M = SS) | | 0.154 | | |
| | | [0.402] | | |
| Treated municipalities (M = SL) | | -0.089 | | |
| | | [0.367] | | |
| Treated municipalities (M = LS) | | 0.215 | | |
| | | [0.435] | | |
| Treated municipalities (M = LL) | | 0.095 | | |
| | | [0.366] | | |
| Treated Ms, post-reform (P x SS = 1) | | -1.099 | | |
| | | [0.339]*** | | |
| Treated Ms, post-reform (P x SL = 1) | | -0.545 | | |
| | | [0.314]* | | |
| Treated Ms, post-reform (P x LS = 1) | | -0.888 | | |
| | | [0.419]** | | |
| Treated Ms, post-reform (P x LL = 1) | | -0.573 | | |
| | | [0.423] | | |
| Population change from reform (in 10,000) | | | -0.01 | 0.013 |
| (Col. 3: Absolute ch., col. 4: Relative ch.) | | | [0.058] | [0.040] |
| Population change x P = 1 | | | -0.158 | -0.103 |
| (Col. 3: Absolute ch, col. 4: Relative ch.) | | | [0.065]** | [0.034]** |
| Vocational training | 0.006 | 0.009 | -0.068 | -0.071 |
| | [0.213] | [0.213] | [0.196] | [0.194] |
| College degree | 0.72 | 0.714 | 0.598 | 0.596 |
| | [0.238]*** | [0.238]*** | [0.226]** | [0.228]* |
| Master's degree or above | 2.161 | 2.153 | 2.07 | 2.074 |
| | [0.366]*** | [0.370]*** | [0.361]*** | [0.363]** |
| Income, 2nd quintile | 0.652 | 0.655 | 0.487 | 0.488 |
| | [0.269]** | [0.266]** | [0.247]* | [0.247] |
| Income, 3rd quintile | 0.861 | 0.858 | 0.696 | 0.686 |
| | [0.311]*** | [0.312]** | [0.299]** | [0.300]* |
| Income, 4th quintile | 2.331 | 2.322 | 2.157 | 2.153 |
| | [0.371]*** | [0.369]*** | [0.348]*** | [0.349]** |
| Income, 5th quintile | 2.911 | 2.908 | 2.89 | 2.906 |
| | [0.475]*** | [0.472]*** | [0.424]*** | [0.421]** |
| Publicly employed | 0.914 | 0.908 | 0.972 | 0.965 |
| | [0.223]*** | [0.224]*** | [0.198]*** | [0.197]** |
| Female | -0.973 | -0.975 | -1.054 | -1.057 |
| | [0.167]*** | [0.171]*** | [0.154]*** | [0.157]** |
| Age | 0.124 | 0.123 | 0.147 | 0.146 |
| | [0.034]*** | [0.033]*** | [0.035]*** | [0.035]** |
| Age^2 | -0.001 | -0.001 | -0.001 | -0.001 |
| | [0.000]** | [0.000]** | [0.000]*** | [0.000]** |
| Observations | 1992 | 1992 | 2251 | 2251 |
| R-squared | 0.15 | 0.15 | 0.15 | 0.15 |
| Sample | Full | Full | Full | Full |
| | sample | sample | sample | sample |

Robust standard errors corrected for clustering at the (new) municipal level in brackets.

* significant at 10%; ** significant at 5%; *** significant at 1%.

A constant was included in all regressions, but is not reported.

**Table 5: Matching DiD-analysis of the effect of municipal amalmagation and size on political efficacy**

| | | No exact treatment group matching | | | | | |
| | | Weighted sample ($n$ = 1026/2152) | | | | | |
| Binary treatment | | Categorical treatment | | Population difference | | Relative pop. difference | |
| ATT_B | -1.028 | ATT_SS | -1.762 | ATT_P | -0.242 *** | ATT_RP | -0.223 *** |
| | [.374] *** | | 0.525 *** | | 0.078 | | 0.058 |
| | | ATT_SL | -1.447 | | | | |
| | | | 0.655 ** | | | | |
| | | ATT_LS | -0.670 | | | | |
| | | | 0.835 | | | | |
| | | ATT_LL | -0.349 | | | | |
| | | | 0.884 | | | | |

| | | Exact treatment group matching on: Pre-reform municipality | | | | | |
| | | Weighted sample ($n$ = 759/1834) | | | | | |
| Binary treatment | | Categorical treatment | | Population difference | | Relative pop. difference | |
| ATT | -1.328 | ATT_SS | -1.940 | ATT_P | 0.269 ** | ATT_RP | -0.248 *** |
| | [.529] ** | | 0.571 *** | | 0.116 | | 0.073 |
| | | ATT_SL | -3.079 | | | | |
| | | | 1.596 * | | | | |
| | | ATT_LS | -0.371 | | | | |
| | | | 0.856 | | | | |
| | | ATT_LL | -0.386 | | | | |
| | | | 0.557 | | | | |

Notes:

Matching-procedure based on two-dimensional Mahalanobi's metric matching with caliper = .025.

Top and bottom panels have, respectively, no exact matching and exact matching within

pre-reform municipalities; see text for further details.

Standard errors, reported in parentheses, are corrected for clustering at the new

  municipal level, bootstrapped with 50 replications and resampling at the cluster level.

**Table 6: Retrospective evaluations of difference in political efficacy before and after reform.**

| | | |
|---|---|---|
| Binary treatment | -1.376 | -1.291 |
| | 0.313 *** | 0.378 *** |
| Categorical treatment | | |
| SS | -2.099 | -2.158 |
| | 0.455 *** | 0.613 *** |
| SL | -2.243 | -1.832 |
| | 0.523 *** | 0.513 *** |
| LS | -1.223 | -0.827 |
| | 0.301 *** | 0.549 |
| LL | -0.472 | -0.394 |
| | 0.272 * | 0.546 |
| Test for coefficient equality§ | p = .000 | p = .063 |
| Population difference, absolute | -0.301 | -0.328 |
| (measured in 10,000s) | 0.054 *** | 0.077 *** |
| Population difference, relative | -0.184 | -0.25 |
| | 0.061 *** | 0.065 *** |
| Matching, binary treatment | -1.555 | -1.307 |
| | 0.348 *** | 0.379 *** |
| Bias corrected c.i. | [-.646;-2.171] | [-.473;-1.831] |
| Sample | Full sample | Weighted matched sample |
| No. of observations | 1031 | 1352 |

§ Reports p-values associated with the test of the null hypothesis that the coefficients
  of the treatment categories are equal against the alternative that they are not equal.
Note: All standard errors are bootstrapped using 200 replications, accounting for
       clustering at the post-reform municipal level.

Figure 1: Distribution of propensity scores across treatment status and time
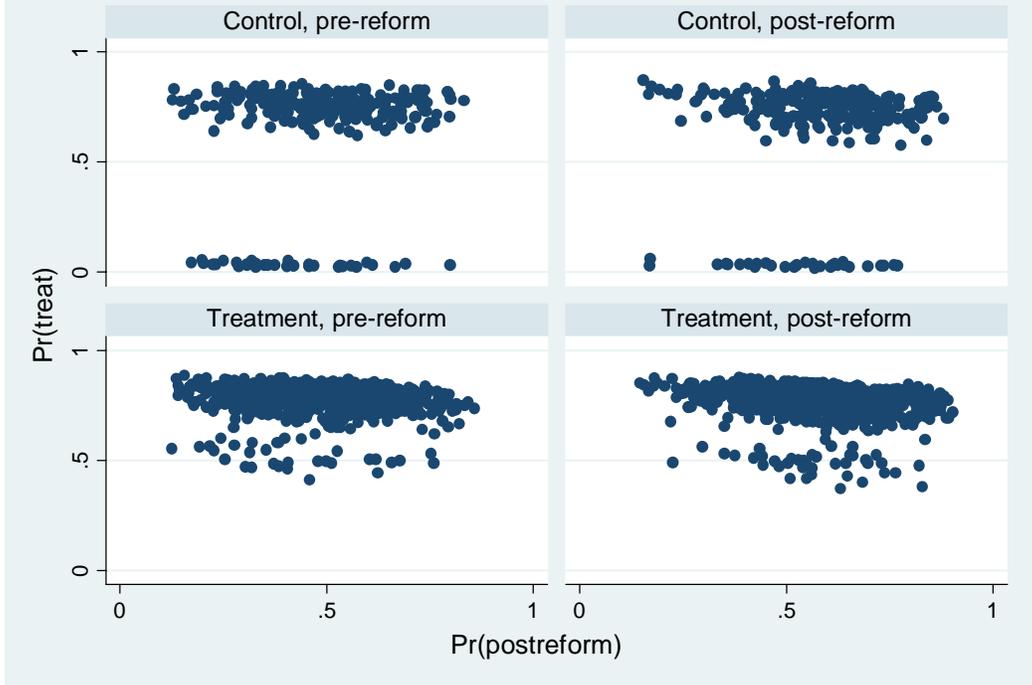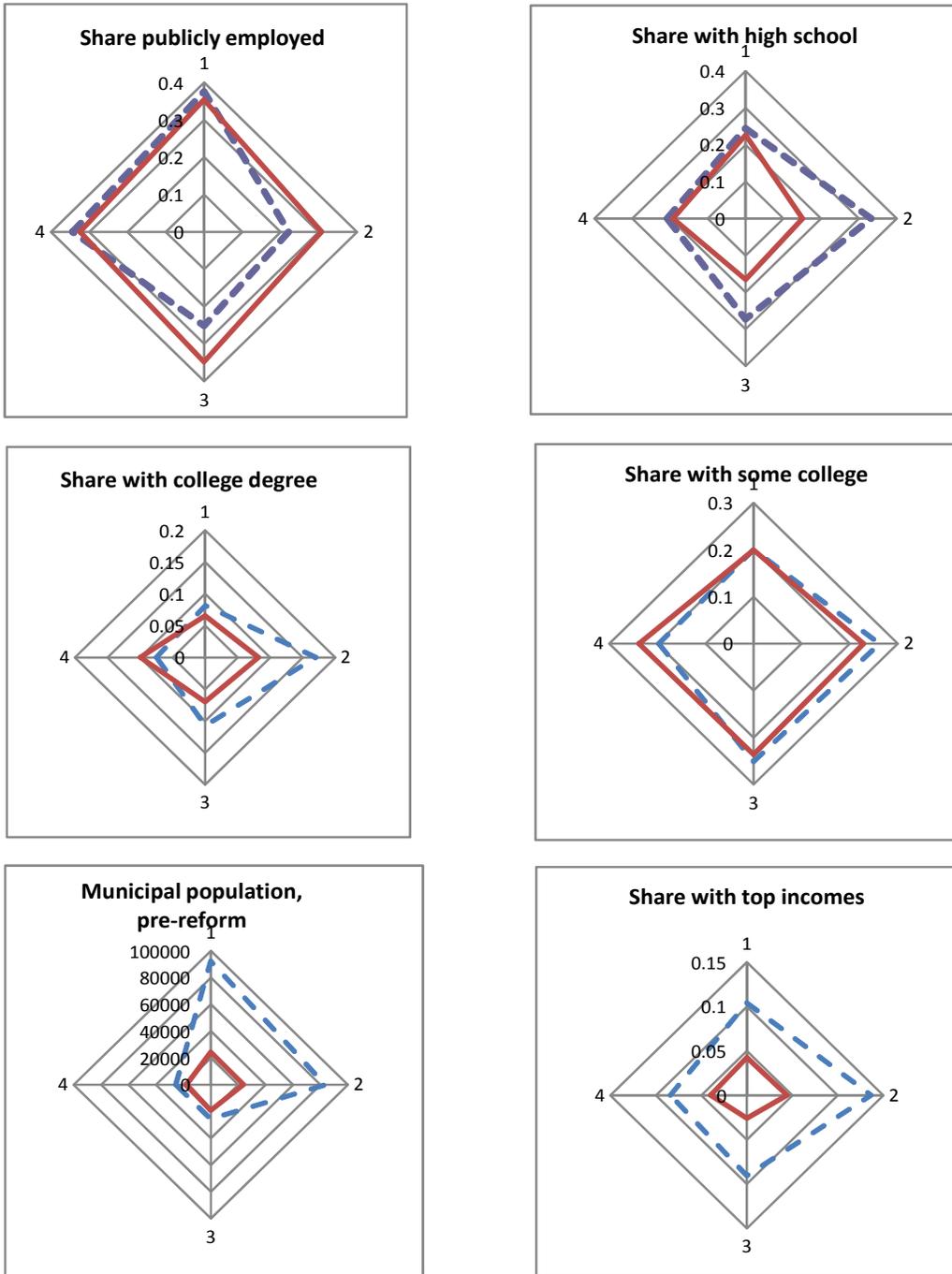
Figure 2: Balance of covariates, before and after matching.
Blue, dashed line: Before matching. Red, solid line: After matching.



**Share publicly employed**

**Share with high school**

**Share with college degree**

**Share with some college**

**Municipal population, pre-reform**

**Share with top incomes**

Note: The closer a box is to being a square, the closer are average values of the covariates and the better is the match between treatment and control groups.

**Table A: Municipalities in the sample**

| Stratum | Selected municipalities | Size 2006 | Size 2007 | Pop. Growth (percent) |
|---|---|---|---|---|
| Not merged | København | 501,158 | 503,699 | 0.5 |
| Not merged | Dragør* | 13,154 | 13,184 | 0.2 |
| Not merged | Gladsaxe | 61,735 | 61,945 | 0.3 |
| Not merged | Høje Taastrup | 46,257 | 46,683 | 0.9 |
| Not merged | Vallensbæk* | 12,230 | 12,145 | -0.7 |
| Not merged | Helsingør | 61,340 | 61,012 | -0.5 |
| Not merged | Ringsted | 31,094 | 31,468 | 1.2 |
| Not merged | Læsø | 2,091 | 2,058 | -1.6 |
| Small before, now small part | Ramsø* | 9,412 | 81,017 | 760.8 |
| Small before, now small part | Holmegaard* | 7,643 | 80,133 | 948.4 |
| Small before, now small part | Glamsbjerg | 5,924 | 41,816 | 605.9 |
| Small before, now small part | Augustenborg | 6,525 | 76,825 | 1077.4 |
| Small before, now small part | Gram | 4,867 | 56,275 | 1056.3 |
| Small before, now small part | Lundtoft | 6,150 | 60,044 | 876.3 |
| Small before, now large part | Ledøje-Smørum | 10,797 | 40,057 | 271.0 |
| Small before, now large part | Jægerspris | 9,520 | 43,910 | 361.2 |
| Small before, now large part | Stenlille | 5,634 | 28,956 | 414.0 |
| Small before, now large part | Trundholm | 11,311 | 32,980 | 191.6 |
| Small before, now large part | Ejby | 10,192 | 36,771 | 260.8 |
| Small before, now large part | Hadsten | 11,969 | 45,037 | 276.3 |
| Large before, now small part | Frederikssund | 19,144 | 43,910 | 129.4 |
| Large before, now small part | Korsør* | 20,873 | 76,949 | 268.7 |
| Large before, now small part | Hedensted | 17,190 | 44,892 | 161.2 |
| Large before, now small part | Grenaa | 18,673 | 38,333 | 105.3 |
| Large before, now small part | Viborg | 44,505 | 91,405 | 105.4 |
| Large before, now small part | Hjørring | 35,296 | 67,118 | 90.2 |
| Large before, now large part | Hillerød | 38,102 | 46,354 | 21.7 |
| Large before, now large part | Middelfart* | 20,599 | 36,771 | 78.5 |
| Large before, now large part | Randers | 62,524 | 92,984 | 48.7 |
| Large before, now large part | Silkeborg | 55,906 | 86,540 | 54.8 |
| Large before, now large part | Skive | 27,972 | 48,344 | 72.8 |
| Large before, now large part | Aalborg | 163,952 | 194,149 | 18.4 |